

Vers un système d'information stratégique en téléphonie mobile prépayée

Cédric ARCHAUX (*,**), **Ali KHENCHAF** (**)

carchaux@bouyguestelecom.fr, Cedric.Archaux@ensieta.fr, Ali.Khenchaf@ensieta.fr

(*) Bouygues Telecom, 20 quai du point du jour, 92640 Boulogne Billancourt, France,

(**) Laboratoire E³I², ENSIETA, 2 rue François Verny, 29806 Brest CEDEX 09, France.

Mots clefs :

Téléphonie Mobile, Aide à la décision, Extraction de Connaissances, Apprentissage Supervisé, Machines à Support de Vecteurs, Analyse de Survie

Keywords:

Mobile Telephony, Decision Support, Knowledge Discovery, Supervised Learning, Support Vector Machines, Survival Analysis

Palabras clave :

Telefonía Móvil, Ayuda a la Decisión, Extracción de Conocimientos, Aprendizaje Supervisado, Máquinas de Soporte de Vectores, Análisis de Supervivencia

Résumé :

L'offre de téléphonie mobile actuelle étant très diversifiée en France, la fidélisation des clients est devenue un enjeu majeur pour les opérateurs. Nous présentons dans cet article des travaux qui tendent vers la conception et le déploiement d'un système d'information stratégique dédié à la gestion des clients en téléphonie mobile prépayée. Nos travaux se situent dans le cadre de l'extraction de connaissances à partir de données et portent sur la caractérisation et la prévision des comportements des clients. Après avoir introduit la spécificité de la téléphonie prépayée et les types de données auxquelles nous sommes confrontés, nous décrivons une première étape de détection de 'fuite' des clients. Nous présentons ensuite un modèle de prévision d'activité qui permet d'établir des estimations financières de la valeur des clients. Nous mettons enfin en perspective le système d'information stratégique qui émerge de la fusion des indicateurs fournis par les différents modèles.

1 Introduction

Après plusieurs années de très forte croissance, le marché français de la téléphonie mobile a atteint sa maturité et se stabilise. Il devient désormais primordial pour les opérateurs de fidéliser leurs clients afin de prévenir leur fuite à la concurrence appelée « *churn* » dans la littérature anglophone. Il est par ailleurs souhaitable d'avoir une estimation de l'activité des clients afin de répondre à leurs attentes. Nous sommes ici placés dans le contexte de la téléphonie mobile prépayée qui ne présente aucune forme d'engagement 'direct', contrairement aux forfaits qui engagent contractuellement les clients à consommer tous les mois un volume minimal pendant une durée déterminée à l'avance. Les clients souscrivant à ce type d'offres achètent à l'avance un crédit de consommation qu'ils peuvent consommer quand ils le souhaitent tant qu'il est valide, et peuvent cesser leur activité sans préavis en arrêtant de recharger leurs lignes.

Nous présentons dans un premier temps les bases de données sur lesquelles nous travaillons en décrivant le type de variables utilisées, nous présentons ensuite la détection de la fuite des clients, qui vise à distinguer ceux qui cesseront leur activité dans un horizon de temps donné de ceux qui maintiendront leur niveau de consommation. Nous présentons dans un troisième temps la modélisation du délai de rechargement. Nous présentons ensuite le modèle de valeur économique des clients en enchaînant sur les résultats obtenus par les modèles de rechargement et de détection de fuite sur nos bases de données. Nous concluons enfin sur nos perspectives d'intégration des différents éléments stratégiques dans un système d'information décisionnel dédié à la gestion de la relation client dans le domaine de la téléphonie mobile prépayée.

2 Bases de données

Les bases de données que nous utilisons dans cette étude caractérisent un ensemble de clients utilisant des lignes prépayées. On dispose pour chaque client de différents types d'informations telles que :

- des données de facturation comme les montants rechargés ou les montants prélevés pour les services et options souscrits, ces montants sont des nombres réels qui prennent généralement leurs valeurs dans un ensemble restreint.
- des données relatives aux usages comme le nombre total des appels, la répartition des appels locaux nationaux ou internationaux (pourcentage), la consommation en pic et la consommation moyenne (réels),
- des données relatives à la ligne téléphonique telles que l'ancienneté (entier borné à l'ancienneté de commercialisation des offres prépayées), le plan tarifaire actuel, le nombre de plans tarifaires différents par lequel le client est passé,
- des données relatives aux souscriptions et résiliations de services,
- d'autres informations telles que l'âge ou la catégorie socioprofessionnelle du client, la rentabilité actuelle et la rentabilité précédente, la sélection d'autres options, etc.

Chaque client est ainsi caractérisé par un ensemble de k variables illustratives ou co-variables. Nous notons (x_{i1}, \dots, x_{ik}) le vecteur formé par les variables illustratives décrivant le client i .

3 Détection de la fuite des clients

Les approches déjà étudiées dans le domaine de la téléphonie se sont tournées vers des techniques telles que les chaînes de Markov [14], les mixtures de gaussiennes et réseaux bayésiens [27], les règles d'associations [25], ou encore les réseaux de neurones [22]. Dans [21], il est montré que l'introduction d'un réseau de neurones multi-couches pour la modélisation de la survie permet d'obtenir de bons résultats en s'appuyant sur une analyse critique d'autres méthodes.

Nous présentons donc l'application de cette approche pour la détection du churn. Cependant, nous avons souhaité améliorer la capacité de généralisation du modèle pour tenir compte du volume très important de données auquel nous sommes confrontés et présentons ainsi la détection par machines à support de vecteurs.

L'approche SVM [28] tente de séparer dans l'ensemble des clients ceux qui présentent un fort risque de fuite des autres moins risqués par l'hyperplan optimal qui garantit que l'écart entre les deux classes soit maximal. Les nouveaux clients pour lesquels nous devons détecter le churn, pourront ainsi ne pas être trop similaires à ceux employés pour trouver l'hyperplan mais être tout de même situés franchement d'un côté ou l'autre de la frontière. La force des SVM tient à leur simplicité de mise en œuvre face à des problèmes difficiles et à des fondements mathématiques solides. Nous avons donc retenu et testé les modèles SVM qui ont fait leur preuves dans d'autres domaines tels que la classification d'images [10], [19], ou la reconnaissance de locuteurs [13]. Nous comparons ainsi la détection par perceptron multicouches à la détection par SVM.

Nous présentons dans cette partie comment la méthode SVM s'inscrit dans le cadre de la théorie de l'apprentissage supervisé et comment nous pouvons formaliser le problème de détection de churn à l'aide de cette approche. Nous rappelons enfin le principe des SVM.

3.1 Théorie d'apprentissage supervisé

Soit O un ensemble de clients décrit par un nombre fixe d de caractéristiques (variables descriptives numériques). Prenons un sous-ensemble S de O , l'ensemble de test constitué d'un ensemble de l couples $(x_i, y_i)_{1 \leq i \leq l}$ où x_i est un point de \square^d qui représente les caractéristiques des clients et $y_i = \pm 1$ représente la classe du client x_i (le client est à risque $(+1)$ ou non (-1)). Etant donné les caractéristiques des clients de $O - S$, l'ensemble d'apprentissage, nous cherchons à estimer si un client de S est à risque ou non, ou encore une estimation de la fonction qui à tout x_i associe un y_i pour l'appliquer à ce nouveau client. Nous cherchons donc la fonction qui réalise la meilleure approximation de la réponse désirée parmi une famille de fonctions $\{f_\alpha\}$ à valeurs dans $\{-1, +1\}$.

Les $(x_i, y_i)_{1 \leq i \leq l}$, supposées indépendantes et identiquement distribuées, sont issues d'une distribution de probabilité inconnue $P(x, y)$. Le critère choisi est la minimisation du risque R défini par :

$$R(\alpha) = \int 1/2 |y - f_\alpha(x)| dP(x, y)$$

La probabilité P étant inconnue, R l'est aussi ; par contre, nous pouvons estimer risque empirique sur l'ensemble des observations de la base d'apprentissage :

$$R_{emp}(\alpha) = 1/(2l) \sum_{i=1}^l |y_i - f_\alpha(x_i)|$$

Pour une probabilité au moins égale à $1 - \eta$, on a l'inégalité suivante :

$$R(\alpha) \leq R_{emp}(\alpha) + \sqrt{1/l(h(\ln(2l/h) + 1) - \ln(\eta/4))} \quad (1)$$

où h est la VC-dimension du nom de Vapnik et de Chervonenkis [11], c'est le maximum de points pour lesquels les fonctions $\{f_\alpha\}$ associent la bonne classe. Le second terme de la borne supérieure, nommé intervalle de confiance, est une fonction croissante monotone en h . Ainsi, pour h petit, il suffit de minimiser le risque empirique pour minimiser le risque R .

Ainsi, pour garantir une faible valeur de R , nous devons chercher une valeur optimale de la VC-dimension h . C'est un problème de minimisation du risque. Le contrôle du risque consiste donc à contrôler la VC-dimension puisque la taille de l'observation l est généralement fixée. Vapnik [28] propose d'appliquer le principe de minimisation du risque structurel dont le but est la minimisation conjointe du risque empirique et de l'intervalle de confiance. En considérant les hyperplans sur \square^d définis par $\{x \in \square^d : x \cdot w + b = 0\}$, [6] montre que minimiser la VC-dimension revient à minimiser $\|w\|$.

3.2 Principe des SVM

S'il existe un hyperplan qui sépare les deux classes, les points de l'hyperplan sont décrits par l'équation $x_i \cdot w + b = 0$ où w est la normale au plan et $|b|/\|w\|$ la distance entre l'hyperplan et l'origine, voir Fig. 1.

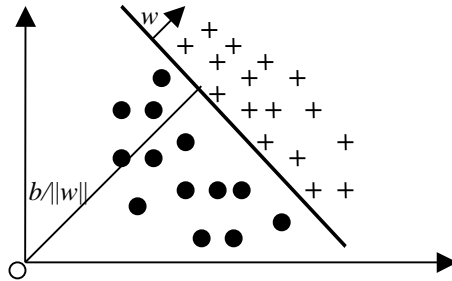


Fig. 1 – Cas linéairement séparable.

Soient d_+ (resp. d_-) la distance minimale entre l'hyperplan et la classe des x_i tel que $y_i = +1$ (resp. $y_i = -1$). L'hyperplan optimal est celui qui maximise $d_+ + d_- = (1-b)/\|w\| - (-1-b)/\|w\| = 2/\|w\|$. Ceci se traduit par l'existence d'un couple $(w, b) \in \mathbb{R}^d \times \mathbb{R}$ tel que : $x_i \cdot w + b = 0$, pour les points de cet hyperplan, avec

$$y_i(x_i \cdot w + b) - 1 \geq 0, \text{ pour tout } i = 1, \dots, l. \quad (2)$$

L'hyperplan optimal est donc déterminé en minimisant $J(w) = \|w\|^2 / 2$ sous les contraintes (2). Les vecteurs de support sont les points tels que $y_i(x_i \cdot w + b) - 1 = 0$. Il s'agit donc de chercher des constantes w et b vérifiant (2) qui minimisent $J(w)$. Ce système se résout simplement [19], et montre que pour estimer la classe d'un nouveau client x , on calcule :

$$f(x) = \text{sign}(x \cdot w^0 + b^0) = \text{sign}\left(\sum_{i=1}^l \alpha_i^0 y_i(x_i, x) + b^0\right) = \text{sign}\left(\sum_{VS} \alpha_i^0 y_i(x_i, x) + b^0\right) \quad (3)$$

où VS est l'ensemble des vecteurs de support.

Pour généraliser cette méthode dans le cas où la fonction de décision n'est pas linéaire, l'idée est de plonger les vecteurs d'entrée dans un autre espace de dimension suffisamment grande en utilisant une fonction $\Phi: \mathbb{R}^d \rightarrow H$, tel qu'il existe une fonction K , le noyau :

$$K: \quad \mathbb{R}^d \times \mathbb{R}^d \Rightarrow \mathbb{R} \\ (x, x') \rightarrow \phi(x) \cdot \phi(x')$$

Il suffit donc de chercher l'hyperplan optimal dans l'espace H par la méthode précédente : le couple $(x_i, y_i)_{1 \leq i \leq l}$ est remplacé par $(\Phi(x_i), y_i)_{1 \leq i \leq l}$, en reprenant les formules précédentes, et en utilisant le produit scalaire de H au lieu du produit scalaire de \mathbb{R}^d . Enfin, pour estimer la classe d'un client x , il suffit de calculer la fonction :

$$f(x) = \text{sign}\left(\sum_{VS} \alpha_i^0 y_i K(x_i, x) + b^0\right)$$

Il n'existe cependant pas de méthode ni pour choisir ϕ , ni pour choisir le noyau K . Les principaux noyaux utilisés dans les applications sont : les polynômes de degré p : $K(x, y) = ((x \cdot y) + 1)^p$, et les gaussiennes $K(x, y) = e^{-\|x-y\|^2 / 2\sigma^2}$.

4 Modélisation du délai de rechargement

Nous cherchons ici à attribuer chaque client un délai de rechargement qui correspond au nombre de jours séparant le dernier rechargement connu du prochain rechargement. Cette attribution d'un délai à un client connaissant ses co-variables est grandement similaires aux études de survie qui visent à attribuer à un patient un délai estimé de survie grâce à l'analyse de ses 'caractéristiques'. Nous suivons donc une approche d'analyse de survie au sens médical du terme.

Les modèles de survie historiquement utilisés sont des modèles paramétriques et des modèles non-paramétriques à risques proportionnels [26]. La modélisation de la survie a été initialisée par les travaux de Kaplan-Meier [18] puis de Cox [9]. La méthode que nous utilisons pour prévoir les rechargements consiste en deux étapes [17], [20]. Une première étape consiste à modéliser le délai de rechargement de chaque client, à partir des rechargements effectués durant une période d'apprentissage. Nous effectuons ensuite une analyse de survie, qui permet d'affecter à chaque client sa probabilité de rechargement dans les six mois suivant la date de modélisation. Certains rechargements ne sont pas observés durant la période d'apprentissage, on dit que les délais qui caractérisent ces rechargements non observés sont censurés.

Si nous appelons T , la variable aléatoire correspondant au délai de rechargement (la donnée de survie), et C la période de temps observée, nous définissons les trois types de censures par :

- Censure à droite : $T > C$
- Censure à gauche : $T < C$
- Censure par intervalle : $C_1 < T < C_2$

Nous sommes dans cette étude confrontés à des données censurées à droite.

4.1 Estimation des délais de rechargement

Nous modélisons les délais de rechargement en ajustant une loi paramétrique à la variable aléatoire décrivant la donnée de survie, la modélisation paramétrique est présentée dans [17]. Nous cherchons ici un modèle qui soit le plus hétérogène possible, c'est à dire qui tienne compte de l'influence des co-variables.

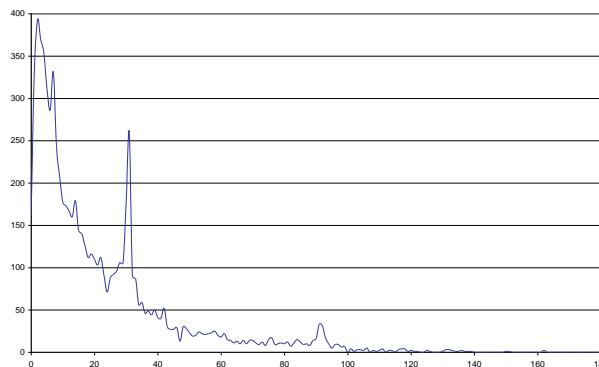


Figure 2 : Graphe de T

Le tracé de la variable T nous amène à penser que cette variable aléatoire suit une loi exponentielle, nous calculons donc le logarithme de cette fonction.

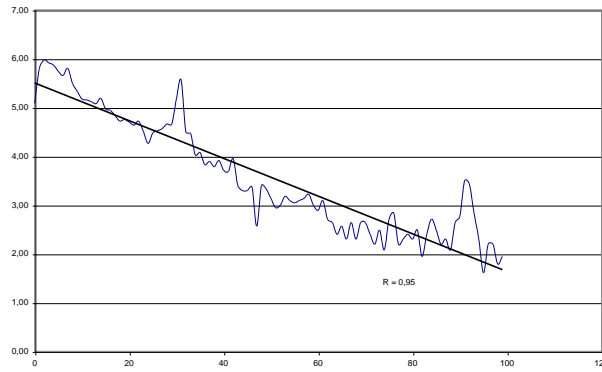


Figure 1 : Graphe de $\ln(T)$

Le logarithme de T est pratiquement linéaire avec un coefficient de corrélation de 95%. Si nous appelons T_i la variable aléatoire caractérisant le délai de rechargement pour un individu i de l'échantillon de travail et x_{i1}, \dots, x_{ik} les valeurs de ses k covariables, on définit :

$$\log T_i = \beta_0 + \beta_1 x_{i1} + \dots + \beta_k x_{ik} + \sigma \varepsilon_i$$

Où ε_i est un terme d'erreur aléatoire et β_0, \dots, β_k et σ sont des paramètres à estimer.

Par passage à l'exponentielle, on obtient une version alternative du modèle :

$$T_i = \exp\{\beta_0 + \beta_1 x_{i1} + \dots + \beta_k x_{ik} + \sigma \varepsilon_i\}$$

Les deux différences entre le modèle et une régression linéaire classique sont

- le terme σ devant le ε et le passage au logarithme de la variable dépendante.
- Le paramètre σ peut être omis, mais cela implique que la variance de ε peut varier d'un jeu de données à un autre. Il est plus simple de fixer la variance de ε à une valeur standard (par exemple 1) et laisser varier le paramètre σ pour compenser la variance du terme d'erreur.

S'il n'y a pas de variable censurée, le modèle peut être facilement estimé par la méthode des moindres carrés, mais nous sommes ici confrontés à des données censurées à droite, c'est pourquoi nous utilisons la méthode du maximum de vraisemblance.

4.2 Estimation des paramètres

Nous avons n individus indépendants ($i=1, \dots, n$), chaque individu i étant caractérisé par ses k covariables x_{i1}, \dots, x_{ik} . Nous pouvons donc caractériser chaque individu par t_i , δ_i et \mathbf{x}_i où t_i est le délai de rechargement de l'individu, δ_i est une variable qui vaut 1 si t_i n'est pas censuré et 0 si t_i est censuré, et \mathbf{x}_i est le vecteur de covariables $\mathbf{x}_i = [x_{i1} \dots x_{ik}]$. Dans le cas où il n'y a pas d'information censurée, la vraisemblance s'exprime sous la forme :

$$L = \prod_{i=1}^n f_i(t_i)$$

Si un individu i est censuré à la date t_i , tout ce que l'on peut dire c'est que son délai de rechargement est au moins égal à t_i . La probabilité que le délai de rechargement soit plus grand que t_i est donnée par la fonction de survie $S(t)$ évaluée à t_i . Si nous avons r observations non censurées et $n-r$ observations censurées, nous pouvons écrire la vraisemblance sous la forme :

$$L = \prod_{i=1}^r f_i(t_i) \prod_{i=r+1}^n S_i(t_i)$$

En utilisant l'indicateur de censure δ nous pouvons écrire la vraisemblance sous une forme équivalente :

$$L = \prod_{i=1}^n [f_i(t_i)]^{\delta_i} [S_i(t_i)]^{1-\delta_i}$$

Lorsque nous choisissons un modèle, nous pouvons remplacer la densité de probabilité et la fonction de survie par leurs expressions respectives. Dans le cas où le modèle choisi est le modèle exponentiel, nous obtenons : $f_i(t_i) = \lambda_i e^{-\lambda_i t_i}$ et $S_i(t_i) = e^{-\lambda_i t_i}$ où $\lambda_i = \exp\{-\beta \mathbf{x}_i\}$ où β_i est un vecteur de coefficients. En remplaçant, on obtient :

$$L = \prod_{i=1}^n [\lambda_i e^{-\lambda_i t_i}]^{\delta_i} [e^{-\lambda_i t_i}]^{1-\delta_i} = \prod_{i=1}^n \lambda_i^{\delta_i} e^{-\lambda_i t_i}$$

Afin de simplifier les calculs, nous passons cette formule au logarithme, ce qui permet de convertir les produits en sommes et les puissances en coefficients, comme la fonction logarithme est croissante, ce qui maximise le logarithme maximise également la fonction :

$$\log L = \sum_{i=1}^n \delta_i \log \lambda_i - \sum_{i=1}^n \lambda_i t_i = -\beta \sum_{i=1}^n \delta_i \mathbf{x}_i - \sum_{i=1}^n t_i e^{-\beta \mathbf{x}_i}$$

Nous maximisons la fonction par la méthode du maximum de vraisemblance.

4.3 Estimation de la probabilité de rechargement

Nous souhaitons $\hat{S}(t)$ de la fonction de survie $S(t) = \text{Prob}(T \geq t)$, qui correspond à la probabilité qu'un rechargement ait lieu après une date donnée. Nous modélisons le délai de rechargement observé par l'estimateur de Kaplan-Meier présenté dans [18] qui est une méthode d'estimation des modèles non paramétriques de données de survie qui permet de tenir compte des données censurées.

Pour une population de clients caractérisée par un rechargement à $t = 0$ date d'origine de la période d'apprentissage, nous mesurons les délais de rechargement, nous les classons en ordre croissant et définissons les quantités suivantes :

- d_j le nombre de rechargements à t_j ,
- n_j le nombre de clients à risque à t_j ,
- c_j le nombre de clients censurés entre t_j et t_{j+1} .

L'estimation de la fonction de survie se fait par l'estimateur de Kaplan Meier :

$$\hat{S}(t) = \prod_{j: t_j \leq t} \left(\frac{n_j - d_j}{n_j} \right)$$

5 Modélisation de la valeur des clients

Le modèle de valeur des clients que nous utilisons est le modèle de valeur à terme LTV (Lifetime Value). Ce modèle a beaucoup évolué ces dernières années comme le témoignent [3], [4], [5], [8], [12], [16], [24] la définissent comme la différence entre les bénéfices prévus pour le client tant qu'il reste client de la société, et les coûts de gestion qui lui sont associés. [23] introduit la notion de valorisation du contact avec les clients. Les références [15] et [7] apportent une analyse plus approfondies des nuances entre les définitions proposées par les références citées ci-dessus.

Nous estimons dans cette étude la valeur à terme des clients, par la définition proposée dans [26]. L'indicateur LTV est composé de la durée de vie d'un client qui est décrite par sa fonction de survie $S(t)$, la valeur estimée du client au cours du temps $V(t)$ et un facteur d'actualisation $D(t)$ qui indique combien vaut aujourd'hui un euro gagné à l'instant t . Avec ces trois composantes, nous définissons la LTV par

$$LTV = \int_0^{\infty} S(t)V(t)D(t)dt$$

5.1 Fonction de Survie

Les rechargements des clients entraînent la survie de leurs lignes téléphoniques, ainsi la fonction de survie utilisée est l'estimateur de Kaplan Meier présenté dans le paragraphe 4.3.

5.2 Valeur estimée

Nous estimons ici la valeur des clients à partir de la valeur de leurs consommations et du montant valorisé de leurs communications entrantes. Nous exprimons ainsi la valeur par :

$$V = V_S + V_E.$$

Où V_S désigne la valeur des consommations et V_E la valeur des communications entrantes. Nous estimons la valeur entrante V_E des clients par régression de leur valeur entrante constatée durant la période d'apprentissage, et nous estimons la valeur V_S par les rechargements effectués, pondérée par la probabilité de rechargement estimée au chapitre 4. Nous développons des modèles spécifiques pour chaque valeur de rechargement, nous lions ainsi la probabilité de valeur à la probabilité de rechargement.

5.3 Fonction d'Actualisation

La valeur de la fonction d'actualisation considérée est une constante obtenue auprès d'experts financiers, il s'agit donc ici d'intégration de connaissance extérieure.

6 Résultats

6.1 Résultats SVM

Nous avons fait apprendre les SVM et le réseau de neurones multicouches sur une même base d'apprentissage de 6000 lignes [2]. Nous appliquons les modèles sur deux bases de tests afin de comparer les résultats. La base de test n°1, constituée de 6000 lignes, vise à tester les modèles sur une base de volumétrie similaire à la base d'apprentissage. La base de test n° 2 de 60.000 lignes vise, elle, à tester les capacités de généralisation des modèles et leur robustesse sur des volumétries significativement supérieures. Le tableau 1 consigne les taux de bonnes classifications des modèles appliqués sur les 2 bases de tests :

Taux de bonne prédiction par modèle	SVM noyau linéaire	SVM noyau gaussien	Réseau de neurones
Test n°1	87,10 %	88,55 %	87,46 %
Test n°2	80,16 %	84,28 %	81,08 %

Tableau 1 : résultats des classificateurs

Nous voyons que sur la base de test n°1 les noyaux gaussiens fournissent des résultats comparables à ceux obtenus par le réseau de neurones et le noyau linéaire. En formulant l'hypothèse que l'estimateur suit une loi gaussienne, nous calculons l'intervalle de confiance à 95% du résultat obtenu par le noyau gaussien sur la base de test n°1 : [86,00 ; 91,10], le taux de bonne prédiction du réseau de neurones et du noyau linéaire ne sont pas significativement plus mauvais.

Le test sur la base n°2 montre que les résultats des SVM sont meilleurs quand la base de test est d'un volume considérablement supérieur à la base d'apprentissage, ce qui est une propriété que nous recherchons. En effet, le taux de bonne prédiction de churn du réseau de neurones et du SVM à noyau linéaire sont à l'extérieur de l'intervalle de confiance à 95% du noyau gaussien [82,85 ; 85,71].

6.2 Résultats Survie

Afin d'optimiser l'ajustement des modèles en tenant compte du nombre de rechargements effectués par les clients durant la période d'apprentissage, nous avons construit trois modèles de survie :

- Le modèle n° 1 pour les clients ayant rechargé une fois,
- Le modèle n° 2 pour les clients ayant rechargé deux fois,
- Le modèle n°3 pour les clients ayant rechargé au moins trois fois,

Nous obtenons ainsi trois modèles de probabilités de rechargement [1].

Nous évaluons la qualité des résultats en comparant les délais de rechargement prévus par le modèle de rechargement au paragraphe 4.1 aux délais que nous constatons réellement sur la base de test. Nous consignons dans le tableau suivant les taux de bonnes prédictions des trois modèles ainsi que les pourcentages de données censurées sur lesquels ils sont construits :

Modèle	Taux de données censurées	Taux de bonnes prédictions
1 rechargement	5,16%	79,93%
2 rechargements	6,63%	79,17%
3 rechargements	6,72%	85,32%

Tableau 2 : Qualité des résultats obtenus

Comme l'indique le tableau, le taux de bonnes prédictions est appréciable, cependant on remarque que les clients rechargeant le moins souvent ont un comportement plus difficile à prévoir, ce qui justifie la détection de la fuite des clients dans une étape séparée.

7 Conclusions et perspectives

Les méthodes présentées dans cet article nous fournissent des résultats satisfaisants et encourageants sur les deux axes de la détection de la propension à fuir et de la probabilité de rechargement des clients.

L'intégration du modèle LTV nous permet d'avoir une estimation de la valeur de l'ensemble des clients et permet d'aller examiner les contributions individuelles pour une exploitation décisionnelle dans le cadre d'actions marketing. Nos perspectives portent sur l'intégration des indicateurs obtenus dans un système d'information décisionnel intégré, allant de la sélection des données jusqu'à la mise à disposition des indicateurs aux utilisateurs finaux du système.

Nous visons ainsi la construction d'indicateurs synthétiques permettant d'avoir une vision instantanée sur les tendances d'évolution de l'ensemble des clients, pour faciliter la prise de décision au niveau macroscopique. Nous souhaitons par ailleurs intégrer les estimations de valeur et de propension à la fuite de chaque client dans les outils de gestion de la clientèle pour faciliter les prises de décision les concernant au cas par cas.

8 Bibliographie

- [1] ARCHAUX C., KHENCHAF A., *Modélisation de la valeur client en téléphonie mobile prépayée*, Sciences of Electronic Technologies of Information and Telecommunications (SETIT), Sousse, Tunisia, 15-20 March 2004
- [2] ARCHAUX C., MARTIN A., KHENCHAF A., *Détection par SVM - Application à la détection de churn en téléphonie mobile prépayée*, Extraction et Gestion des Connaissances (EGC), in *Revue des Nouvelles Technologies de l'Information*, Vol 2, pp. 597, Clermont Ferrand, France, 20-23 Janvier 2004
- [3] BERGER P.D., NASR N.I., *Customer lifetime value: marketing models and applications*, *Journal of Interactive Marketing*, vol.12, n°1, pp. 17-30, 1998
- [4] BITRAN G.R., MONDSCHHEIN S., *Mailing decisions in the catalog sales industry*, *Management Science*, vol. 42, n°9, pp. 1364-1381, 1996
- [5] BLATTBERG R.C., DEIGHTON J., *Manage marketing by the customer equity test*, *Harvard Business Review*, Jul-Aug, pp. 136-144, 1996
- [6] BURGESS C., *A Tutorial on Support Vector Machines for Pattern Recognition*, *Data Mining and Knowledge Discovery*, Vol. 2(2), pp. 121-167, 1998

- [7] CALCIU M., SALERNO F., *Modélisation de la valeur client (Lifetime Value) : synthèse des modèles et propositions d'extension*, 18ème Congrès International de l'Association Française du Marketing, pp. 205-230, 2002
- [8] COURTHEOUX R., *Customer retention: how much to invest*, Research and the Customer Lifecycle, New York, 1995
- [9] COX D.R., *Regression models and life tables*, Journal of the Royal Statistical Society, B34: pp. 187- 220, 1972
- [10] GOFFINET J., *Machines à vecteur de support pour la détection et le suivi de personnes sur des séquences vidéo*, rapport de stage, 2001
- [11] GUERMEUR Y., PAUGAM-MOISY H., *Théorie de l'apprentissage de Vapnik et SVM, Support Vector Machines*, Apprentissage automatique, Hermes Sciences Publications, Paris 1999
- [12] GUPTA S., LEHMANN D.R., *Customers as assets. Journal of Interactive Marketing*, vol. 17, n°1, pp. 9–24, 2003
- [13] GUTSCHOVEN B., VERLINDE P., *Multi-modal Identity Verification using Support Vector Machines (SVM)*, International Conference on Information Fusion, Paris, France, 10-13 juillet 2000
- [14] HOLLMÉN J., *User Profiling and Classification for Fraud Detection*, Thèse de doctorat, University of Helsinki, 2000
- [15] HWANG H.S., JUNG T.S., SUH E.H., *An LTV model and customer segmentation based on customer value*, Expert systems with applications, vol. 26, pp.181-188, 2004
- [16] JACKSON D.R., *Strategic application of customer lifetime value in the direct marketing environment*, Journal of Targeting Measurement and Analysis for Marketing, vol.3, n°1, pp. 9–17, 1994
- [17] KALBFLEISCH J.D., PRENTICE R.L., *The Statistical Analysis of Failure Time Data*, John Wiley & Sons; 2nd edition, ISBN: 047136357X, 2002
- [18] KAPLAN E.L., MEIER R., *Nonparametric Estimation From Incomplete Observations*, Journal of the American Statistical Association, pp. 457-481, 1958
- [19] LAAYANA H., *Détection par SVM – Application à la détection de roches pour le recalage d'images sonar*, rapport de DESA, juillet 2003
- [20] LAWLESS J.F., *Statistical Models and Methods for Lifetime Data*, John Wiley & Sons; 2nd edition, ISBN: 0471372153, 2002
- [21] MANI D.R., DREW J., BETZ A., DATTA P., *Statistics and data mining techniques for lifetime value modelling*, Proceedings of the fifth ACM SIGKDD international conference on Knowledge discovery and data mining, pp. 94-103, 1999
- [22] MOZER M.C., DODIER R., COLAGROSSO M.D., GUERRA-SALCEDO C., WOLNIEWICZ R., *Prodding the ROC Curve : Constrained Optimization of Classifier Performance*, Advances in Neural Information Processing Systems 14, MIT Press, 2002
- [23] PEARSON S., *Building brands directly: creating business value from customer relationships*, London: MacMillan Business, 1996.
- [24] ROBERTS M.L., BERGER P.D., *Direct marketing management*, Prentice-Hall: Englewood Cliffs, 1989
- [25] ROSSET S., MURAD U., NEUMANN E., IDAN Y., PINKAS G., *Discovery of fraud rules for telecommunications-challenges and solutions*, Proceedings ACM SIGKDD, 1999
- [26] ROSSET S., NEUMANN E., EICK U., VATNIK N., IDAN Y., *Customer lifetime value modeling and its use for customer retention planning*, Proceedings of the eighth ACM SIGKDD, pp. 332-340, 2002
- [27] TANIGUCHI M., HAFT M., HOLLMÉN J., TRESP V., *Fraud detection in communications networks using neural and probabilistic methods*, ICCASP, Vol. 2, pp. 1241-1244, 1998
- [28] VAPNIK V., *Statistical Learning Theory*, John Wiley & Sons, 1998